第 1 卷第 2 期
2 0 2 5年 3月

人 文 与 社 会 科 学 学 刊
ISSN 3079-4552 eISSN 3079-4560

Vol. 1 No. 2
Mar. 2 0 2 5

Doi：doi.org/10.70693/rwsk.v1i2.552

# Signal-Based Trading Strategy for SPY ETF: A Multiple Linear Regression Approach

**Zhen Zou**[1]

[1]*Dongbei University of Finance and Economics, Dalian (116025), China*

**Abstract:** This study develops a signal-based trading strategy for the SPDR S&P 500 ETF Trust (SPY) using a multiple linear regression framework to analyze interrelationships between SPY and global equity indices across U.S., European, Asian, and Australian markets. By synthesizing historical pricing data from these major benchmarks, the model generates systematic trading signals through predicted price trajectories. In controlled training scenarios, the strategy achieved superior risk-adjusted returns compared to passive buy-and-hold approaches, demonstrating the value of cross-market signal integration. While the framework shows promise for algorithmic trading systems, the study acknowledges limitations in generalizing historical patterns to evolving market conditions. The findings highlight opportunities to enhance predictive accuracy through machine learning architectures capable of processing nonlinear market dynamics. These insights advance quantitative trading research by establishing methodologies for cross-market signal synthesis and proposing pathways to develop adaptive models for volatile capital markets.

**Keywords:** SPY ETF; Signal-based trading strategy; Multiple linear regression; Global stock indices

## 1. Introduction

Exchange-Traded Funds (ETFs) are a unique and versatile financial instrument that provide investors with access to a diversified basket of assets, such as stocks, bonds, or commodities, without the need to purchase each component individually.

Unlike traditional mutual funds, ETFs trade on exchanges throughout the day at market prices, offering investors greater flexibility and control over their trades. Their ability to provide instant diversification, sectoral exposure, and access to both traditional and alternative investments, including exotic options, further enhances their appeal as a cost-effective and efficient investment vehicle.

Several key factors influence SPY's price movements, including market correlations, macroeconomic conditions, volatility, and global market sentiment, which collectively affect the dynamics of the S&P 500 Index.

To address this, this essay will center on following key areas: the selection of relevant global indices as predictors, the robustness of the regression model in handling financial time-series data, and the performance metrics for evaluating the predictive accuracy and profitability of the strategy. By leveraging the strengths of regression analysis within the Python programming environment, this study aims to produce actionable insights and demonstrate a systematic approach to ETF-based trading strategies.

## 2. Literature Review

### 2.1. The Features of ETFs and Their Role in Financial Markets

Over the past two decades, ETFs have grown into a critical segment of the financial markets, providing exposure to a variety of sectors, indices, and alternative investments (Marszk & Lechman, 2020). Their role in investment portfolios has become even more pronounced due to their ability to efficiently replicate the performance of benchmark indices, such as the S&P 500.

ETFs like SPY, which track the S&P 500 index, are among the most widely traded financial instruments due to their liquidity, transparency, and role in passive investing strategies (Dias et al., 2024). These attributes make ETFs highly attractive for retail and institutional investors alike. Passive strategies underpinned by ETFs, such as SPY, allow investors to gain broad market exposure while minimizing transaction costs and management fees (Hill et al., 2015).

### 2.2. The Association Among Different Stock Markets

Global stock markets have demonstrated varying degrees of interdependence influenced by economic integration, trade relations, and financial crises. The dynamic relationships among the US, Asian, Australian, and European stock markets provide insights into cross-market influences and diversification opportunities.

**US and Asian/Australian Markets:** Studies highlight significant relationships between the US stock market and Asian markets, with the US market often leading in influence. For instance, Dhanaraj et al. (2013) demonstrated that US market movements significantly impact newly industrialized Asian economies such as Hong Kong, Singapore, and South Korea, especially during crises like the 2008 subprime crisis. Similarly, Shamsuddin and Kim (2003) found long-term integration between the Australian stock market, the US, and Japan, although this integration weakened after the Asian financial crisis.

**US and European Markets:** The interdependence between US and European stock markets is strong, with integration studies often citing the US as a leading influence. For example, Sakthivel and Kamaiah (2012) observed robust co-movements between US indices and major European stock markets, reflecting globalization's impact (Sakthivel & Kamaiah, 2012).

**Asian and European Markets:** The connection between Asian and European markets is also evident but generally less pronounced than US-related linkages. Pan et al. (2014) showed strong co-movement between European stock indices, such as Germany's DAX and the UK’s FTSE, and global leaders like the US market, while revealing weaker ties with Asian markets (Pan et al., 2014).

**Post-Crisis and Pandemic Dynamics:** Crises amplify interdependence among global stock markets. For instance, Jamil et al. (2023) revealed that ASEAN-5 markets moved more closely with US and European markets during the COVID-19 pandemic compared to pre-pandemic periods, signifying contagion effects (Jamil et al., 2023).

### 2.3. Previous Studies on ETF Trading Strategies

Several studies have examined trading strategies for ETFs, with SPY serving as a prominent benchmark due to its popularity and robust liquidity. Bollapragada et al. (2013) compared various forecasting techniques, including regression, exponential smoothing, and ARIMA models, to predict SPY prices. Their findings indicated that multiple regression models achieved superior accuracy with low forecast errors.

Non-linear techniques such as neural networks have also been applied to ETF trading. Vasylieva et al. (2020) demonstrated that while linear models like moving averages provide short-term forecasts, neural network models can capture non-linear dynamics in ETF price movements over longer periods. Furthermore, adaptive evolutionary neural networks have been employed to optimize SPY trading strategies, outperforming traditional benchmarks such as moving averages (Sermpinis et al., 2016).

### 2.4. Regression-Based Models in Financial Forecasting

Regression models, particularly multiple linear regression, have been extensively utilized for financial time-series forecasting due to their simplicity and interpretability. Bollapragada et al. (2013) achieved significant predictive accuracy for SPY prices using regression-based techniques, highlighting their suitability for time-series data.

The importance of evaluating model performance using robust metrics such as RMSE, Adjusted

$R^2$, and Sharpe Ratio has been highlighted in recent studies. Liu (2024) employed a signal-based regression model to predict SPY's price changes. Results showed that the regression model outperformed the passive strategy, with improvements in the Sharpe ratio and reduced maximum drawdown.

Additionally, regression-based models have been used to explore correlations between global stock markets. Umair et al. (2018) demonstrated the robustness of regression models when applied to large datasets from multiple global stock exchanges.

## 3. Data Description

### 3.1 Data Source

The data for this analysis is derived from publicly available stock market index data, specifically focused on the relationship between the performance of the SPY ETF and several major global stock market indices:

**Table 1** Data Source and Description

| Indices | Region | Data Range | Description |
|---|---|---|---|
| SPY ETF | United States | 2010/12/31-2021/09/17 | Tracking the performance of the S&P 500 index, is included to provide insight into the broader US market trends. |
| S&P 500 | United States | 2010/12/31-2021/09/17 | Tracking the performance of 500 large companies in the US. |
| DJI | United States | 2010/12/31-2021/09/17 | Sourced from the US stock market, representing 30 major American companies. |
| Dax | Germany | 2011/01/03-2021/09/17 | Correspondent to the German stock market, represented by the DAX index. |
| Nikkei225 | Japan | 2011/01/04-2021/09/17 | Representing the performance of the Japanese stock market. |
| All Ordinary | Australia | 2011/01/04-2021/09/17 | Representing the performance of the Australian stock market, focusing on the All Ordinaries index. |

For each index, its opening, high, low, closing price, adjusted closing prices, and volume on each transaction day are collected (take SPY ETF as an example):

**Table 2** The First Five Rows in SPY ETF Data

| Date | Open | High | Low | Close | Adj. Close | Volume |
|---|---|---|---|---|---|---|
| 2010/12/31 | 125.5300 | 125.8700 | 125.3300 | 125.7500 | 102.0556 | 91218900 |
| 2011/1/3 | 126.7100 | 127.7000 | 125.7000 | 127.0500 | 103.1106 | 138725200 |
| 2011/1/4 | 127.3300 | 127.3700 | 126.1900 | 126.9800 | 103.0538 | 137409700 |
| 2011/1/5 | 126.5800 | 127.7200 | 126.4600 | 127.6400 | 103.5895 | 133975300 |
| 2011/1/6 | 127.6900 | 127.8300 | 127.0100 | 127.3900 | 103.3866 | 122519000 |

From Table 1, the time horizons of the Dax, Nikkei225 and All Ordinary indices are different from those of the other three indices. So, the next step is to process the data for subsequent analysis.

### 3.2 Data Munging

**（1）Data Integration**

To handle time-zone differences effectively, ensuring that the data aligns with the analysis model, a data frame called *Indexpanel* is created, which serves as the foundation for the trading model.

The first step involves calculating the daily price changes for each index:

$$Price\ Change_{Index,t} = Open_{Index,t} - Open_{Index,t-1} \#(1)$$

Additionally, a lagged value of SPY is created by shifting the values by one day, stored in the *spy_lag1* column, which helps in tracking the previous day's movement. Finally, a column named Price is added, which stores the opening price of the SPY ETF.

**Table 3** The First Five Rows in *Indexpanel*

| Date | SPY | SPY_lag1 | DJI | SP500 | DAXI | Nikkei225 | Aord | Price |
|---|---|---|---|---|---|---|---|---|
| 2010/12/31 | 1.1800 | - | - | - | - | - | - | 125.5300 |
| 2011/1/3 | 0.6200 | 1.1800 | 8.0966 | 0.8600 | - | - | - | 126.7100 |
| 2011/1/4 | -0.7500 | 0.6200 | 93.4707 | 15.3300 | 6.2900 | 45.9092 | -9.0000 | 127.3300 |
| 2011/1/5 | 1.1100 | -0.7500 | 17.7100 | -4.1699 | -13.6000 | -7.1801 | -25.2002 | 126.5800 |
| 2011/1/6 | -0.1300 | 1.1100 | 28.3193 | 7.5100 | -16.6001 | 52.2402 | 7.8999 | 127.6900 |

**（2）Missing Value Processing**

After checking, *Indexpanel* has the following number of missing values inside it:

**Table 4** Number of Missing Values

| Column Name | Number of Missing Values |
|---|---|
| SPY | 1 |
| SPY_lag1 | 1 |
| DJI | 1 |
| SP500 | 1 |
| DAXI | 61 |
| Nikkei225 | 163 |
| Aord | 57 |
| Price | 0 |

To fill the missing values, the last known valid value is carried forward, ensuring that no gaps remain in the data.

**Table 5** Indexpanel (2696×8) After Processing Missing Values

| Date | SPY | SPY_lag1 | DJI | SP500 | DAXI | Nikkei225 | Aord | Price |
|---|---|---|---|---|---|---|---|---|

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 2011/1/4 | -0.7500 | 0.6200 | 93.4707 | 15.3300 | 6.2900 | 45.9092 | -9.0000 | 127.3300 |
| 2011/1/5 | 1.1100 | -0.7500 | 17.7100 | -4.1699 | -13.6000 | -7.1801 | -25.2002 | 126.5800 |
| 2011/1/6 | -0.1300 | 1.1100 | 28.3193 | 7.5100 | -16.6001 | 52.2402 | 7.8999 | 127.6900 |
| 2011/1/7 | -0.9800 | -0.1300 | -20.0693 | -1.8800 | 42.0698 | 34.3203 | -12.3999 | 127.5600 |
| 2011/1/10 | 0.8600 | -0.9800 | -24.5205 | -3.5701 | -71.1899 | 34.3203 | 9.2002 | 126.5800 |

**Table 6** Result After Processing Missing Values

| Column Name | Number of Missing Values |
|---|---|
| SPY | 0 |
| SPY_lag1 | 0 |
| DJI | 0 |
| SP500 | 0 |
| DAXI | 0 |
| Nikkei225 | 0 |
| Aord | 0 |
| Price | 0 |

**(3) Data Splitting**

To evaluate the model’s ability to generalize to new, unseen data, the data is split into training and test sets. By using separate training and testing sets, it helps to ensure that the model’s performance reflects its true ability to generalize to real-world data (Hastie, 2009).

**Table 7**Methods of Data Splitting

| | **Indexpanel** | |
|---|---|---|
| | Train Set (1000×8) | Test Set (1000×8) |
| Row Range | 1001-2000 | 1-1000 |

The dataset is divided using an index-based panel structure, where observations are ordered sequentially. In this study, the first 1000 rows (1-1000) are allocated to the test set, while the subsequent 1000 rows (1001-2000) form the training set. This backward allocation approach ensures that the test data precedes the training data, simulating a real-world scenario where future observations are predicted based on past information.

## 3.3 Training Data Set Exploration

### (1) Descriptive Statistical Analysis

For the training set, the descriptive statistical analysis (use Excel) provides a comprehensive understanding of the central tendency, dispersion, and distributional characteristics of the selected indices and SPY price over the observation period.

**Table 8** Descriptive Statistical Analysis of Training Data

| | **spy** | **spy_lag1** | **dji** | **sp500** | **daxi** | **nikkei** | **aord** | **Price** |
|---|---|---|---|---|---|---|---|---|
| Mean | 0.0464 | 0.1009 | 6.5604 | 0.9946 | 1.7816 | -1.8197 | 0.7221 | 224.6636 |
| Std. Error | 0.0580 | 0.0612 | 5.2416 | 0.5791 | 3.8088 | 5.1672 | 1.3373 | 1.3205 |
| Median | 0.1100 | 0.2000 | 11.5947 | 1.7750 | 8.0898 | 0.8848 | 3.7000 | 210.6200 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Mode | -0.3100 | -0.2500 | 59.6895 | 6.5500 | -22.1201 | -209.7197 | 6.5000 | 207.2900 |
| Std. Dev | 1.8343 | 1.9354 | 165.7529 | 18.3115 | 120.4455 | 163.4010 | 42.2889 | 41.7582 |
| Kurtosis | 8.5388 | 10.0362 | 7.5020 | 6.7085 | 4.2047 | 9.5139 | 2.1315 | 4.6295 |
| Skewness | -0.7985 | -1.3034 | -1.0829 | -1.0512 | -0.7316 | -0.9550 | -0.6515 | 1.6726 |
| Min | -14.2400 | -14.2400 | -1252.6992 | -126.2800 | -884.8496 | -1381.8506 | -204.3999 | 163.5500 |
| Max | 8.8000 | 8.800 | 807.6992 | 76.1699 | 414.2500 | 1012.3789 | 129.5996 | 444.5300 |
| Sum | 46.3797 | 100.9400 | 6560.4463 | 994.6031 | 1781.5552 | -1819.7422 | 722.0981 | 224663.6301 |
| Observation | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 |
| C.I. (95%) | 0.1138 | 0.1201 | 10.2857 | 1.1363 | 7.4742 | 10.1398 | 2.6242 | 2.5913 |

While SPY and SP500 reflect stability, indices like DJI and Nikkei exhibit significant risk and return variability, highlighting the need for investors to consider volatility and distribution characteristics.

### (2) Correlation Analysis

By generating a scatter plot matrix for the *Train* dataset, this process enables an intuitive understanding of the data's structure and inter-variable relationships.

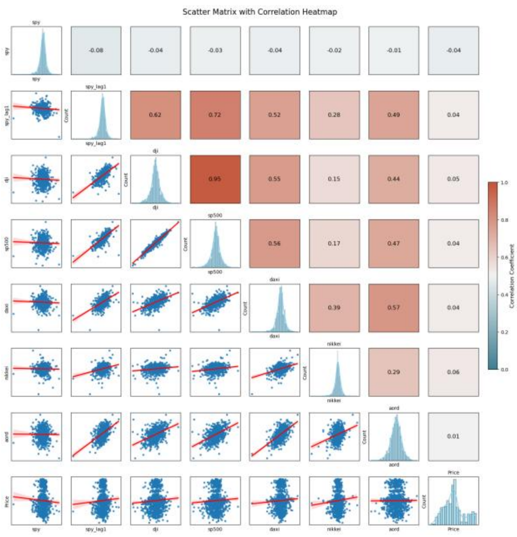


**Figure 1** Scatter Matrix

According to the figure, Strong **positive linear trends** exist between dji, sp500, and spy_lag1, while variables like aord and Price show weak relationships.

The diagonal of the matrix shows the **individual distributions**, most variables follow bell-shaped distributions, indicating normality, with aord exhibiting higher variability.

As to **correlation coefficients**, U.S. indices (dji and sp500) are highly correlated. Weak correlations are observed between global indices (e.g., aord, nikkei) and U.S. markets.

Specifically, the correlation coefficients between each variable and spy are examined further:

**Table 9** Correlation Coefficient with spy Explanation

| **Correlation Coefficient with** | **Explanation** |
|---|---|

| | **spy** | |
|---|---|---|
| spy_lag1 | -0.0800 | The lagged value of SPY (spy_lag1) has a **weak negative correlation** (-0.080) with the current SPY value. This suggests a slight inverse relationship, implying that previous SPY performance has a marginal impact on its current movement, but the relationship is not strong enough to be considered significant. |
| dji | -0.0400 | The Dow Jones Industrial Average (dji) shows a very **weak negative correlation** (-0.040) with SPY. This indicates that movements in the Dow Jones Index have little to no direct linear relationship with the SPY ETF. |
| sp500 | -0.0290 | The S&P 500 Index (sp500) also demonstrates a **very weak negative correlation** (-0.029) with SPY. Although SPY itself tracks the S&P 500, the negative value might reflect short-term differences due to lag effects or market volatility. |
| dax | -0.0370 | The German DAX Index (daxi) exhibits a similarly **weak negative correlation** (-0.037) with SPY. This suggests that the European market (DAX) has minimal influence on SPY's movement, reflecting a lack of immediate connection. |
| nikkei225 | -0.0180 | The Japanese Nikkei Index (nikkei) has a near-zero **negative correlation** (-0.018) with SPY. This indicates virtually no linear relationship between SPY and Japan's stock market performance. |
| aord | -0.0080 | The Australian All Ordinaries Index (aord) has the weakest **negative correlation** (-0.008) with SPY, essentially suggesting no relationship at all. SPY and AORD operate in geographically and temporally separate markets, which may explain the lack of correlation. |

The weak correlation coefficients suggest that predictors individually explain little variation in SPY. However, **multiple linear regression** examines the combined effect of predictors, which may reveal relationships not apparent in pairwise correlations.

## 4. Methodology

### 4.1 Multiple Linear Regression Model Fitting

Multiple linear regression is a fundamental statistical technique used to model the relationship between a single dependent variable and multiple independent (predictor) variables (Tsay, 2010; Hastie, Tibshirani & Friedman, 2009).

In this case, the dependent variable is *spy* (the S&P 500 ETF), and the predictors include spy_lag1, sp500, dji, daxi, aord, and nikkei.

$$\begin{aligned} spy_t = \beta_0 + \beta_1 spy_{t-1} + \beta_2 sp500_t + \beta_3 dji_t + \beta_4 daxi_t \\ + \beta_5 aord_t + \beta_6 nikkei_t + \varepsilon_t \end{aligned} \#(2)$$

To estimate the regression coefficients, write the regression equation in matrix form (Montgomery, Peck & Vining, 2012):

$$\underbrace{\begin{pmatrix} spy_{1001} \\ spy_{1002} \\ \vdots \\ spy_{2000} \end{pmatrix}}_{\boldsymbol{Y}} = \underbrace{\begin{pmatrix} 1 & spy_{lag1_{1001}} & \cdots & nikkei_{1001} \\ 1 & spy_{lag1_{1002}} & \cdots & nikkei_{1002} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & spy_{lag1_{2000}} & \cdots & nikkei_{2000} \end{pmatrix}}_{\boldsymbol{X}} \underbrace{\begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_6 \end{pmatrix}}_{\boldsymbol{\beta}} + \underbrace{\begin{pmatrix} \varepsilon_{1001} \\ \varepsilon_{1002} \\ \vdots \\ \varepsilon_{2000} \end{pmatrix}}_{\boldsymbol{\varepsilon}} \#(3)$$

$$\hat{\beta} = (X^T X)^{-1} X^T Y \#(4)$$

The regression results obtained through Python are as follows[1]:

**Table 10** OLS Regression Results

| OLS Regression Results | | | | | | |
|---|---|---|---|---|---|---|
| **Dep. Variable:** | | spy | | **R-squared:** | | 0.015 |
| **Model:** | | OLS | | **Adj. R-squared:** | | 0.009 |
| **Method:** | Least Squares | | | **F-statistic:** | | 2.448 |
| **Observations:** | | 1000 | | **Prob(F-statistic):** | | 0.0235 |
| **Df Residuals** | | 993 | | **Log-likelihood:** | | -1850.8 |
| **Df Model:** | | 6 | | **AIC:** | | 3716 |
| **Covariance:** | | nonrobust | | **BIC:** | | 3750 |
| | **coef** | **std err** | **t** | **P>\|t\|** | **[0.025** | **0.975]** |
| **Intercept** | 0.0926 | 0.049 | 1.888 | 0.059 | -0.004 | 0.189 |
| **spy_lag1** | -0.1697 | 0.051 | -3.349 | 0.001 | -0.269 | -0.070 |
| **sp500** | 0.0351 | 0.014 | 2.587 | 0.010 | 0.008 | 0.062 |
| **dji** | -0.0032 | 0.001 | -2.280 | 0.023 | -0.006 | 0.000 |
| **daxi** | -0.0002 | 0.001 | -0.382 | 0.702 | -0.001 | 0.001 |
| **aord** | 0.0015 | 0.001 | 1.057 | 0.291 | -0.001 | 0.004 |
| **nickel** | 5.227E-05 | 0.000 | 0.159 | 0.874 | -0.001 | 0.001 |
| **Omnibus** | | 320.467 | | **Durbin-Watson:** | | 1.991 |
| **Prob(Omnibus)** | | 0.000 | | **Jarque-Bera (JB):** | | 4234.114 |
| **Skew** | | -1.083 | | **Prob(JB):** | | 0.000 |
| **Kurtosis** | | 12.845 | | **Cond. No.** | | 194.000 |

**Regression Function:**

$$\begin{aligned} spy_t = \underset{(0.049)}{0.0926} - \underset{(0.051)}{0.1697} spy_{t-1} + \underset{(0.014)}{0.0351} sp500_t - \underset{(0.001)}{0.0032} dji_t \\ - \underset{(0.001)}{0.0002} daxi_t + \underset{(0.001)}{0.0015} aord_t + \underset{(0.000)}{0.00005227} nikkei_t \end{aligned} \#(5)$$

**Model Fit:** $R^2 = 0.015$ means the model explains only 1.5% of the variation in the dependent variable (spy), indicating that most of the variability remains unexplained by the predictors. However, after accounting for the number of predictors, the explanatory power (Adj. $R^2 = 0.009$) of the model remains weak. The F-test indicates that the overall regression is statistically significant at the 5% level. However, the low F-statistic value suggests only a modest relationship between the dependent

1 Standard Errors calculated in the table assume that the covariance matrix of the errors is correctly specified.

and independent variables.

**Coefficients and Statistical Significance:** The intercept is marginally insignificant, implying that when all predictors are zero, the expected value of spy is not definitively different from zero. While all other slope coefficients are significant, daxi (-0.0002, p = 0.702), aord (0.0015, p = 0.291), nikkei (5.227E-05, p = 0.874): These variables are statistically insignificant, meaning their relationship with spy cannot be reliably determined from this model.

**Model Diagnostics: Omnibus** (320.467, p = 0.000) and **Jarque-Bera** (4234.114, p = 0.000) indicate that the residuals deviate significantly from normality. The high **kurtosis** (12.845) and negative **skew** (-1.083) suggest heavy tails and a left-skewed distribution of residuals. **Durbin-Watson** (1.991) is close to 2, indicating no significant autocorrelation in the residuals. **Condition Number** (194), a relatively high value, signals potential multicollinearity among the predictors. Further investigation of the correlation matrix confirms some level of collinearity, particularly with variables like sp500 and dji.

### 4.2 Multicollinearity Test

Multicollinearity occurs when two or more predictor variables are highly correlated, making it difficult to estimate their individual effects on the dependent variable. In this model, multicollinearity is tested in predictors which failed the p-value test (James et al., 2013).

**Table 11** Correlation Matrix

| | **spy** | **spy_lag1** | **dji** | **sp500** | **daxi** | **nikkei** | **aord** | **Price** |
|---|---|---|---|---|---|---|---|---|
| **spy** | 1.0000 | -0.0801 | -0.0402 | -0.0291 | -0.0370 | -0.0184 | -0.0079 | -0.0428 |
| **spy_lag1** | -0.0801 | 1.0000 | 0.6222 | 0.7213 | 0.5234 | 0.2814 | 0.4949 | 0.0427 |
| **dji** | -0.0402 | 0.6222 | 1.0000 | 0.9545 | 0.5468 | 0.1452 | 0.4391 | 0.0546 |
| **sp500** | -0.0291 | 0.7213 | 0.9545 | 1.0000 | 0.5590 | 0.1695 | 0.4726 | 0.0448 |
| **daxi** | -0.0370 | 0.5234 | 0.5468 | 0.5590 | 1.0000 | 0.3928 | 0.5728 | 0.0440 |
| **nikkei** | -0.0184 | 0.2814 | 0.1452 | 0.1695 | 0.3928 | 1.0000 | 0.2893 | 0.0579 |
| **aord** | -0.0079 | 0.4949 | 0.4391 | 0.4726 | 0.5728 | 0.2893 | 1.0000 | 0.0062 |
| **Price** | -0.0428 | 0.0427 | 0.0546 | 0.0448 | 0.0440 | 0.0579 | 0.0062 | 1.0000 |

By examining dji, daxi, nikkei from the correlation matrix, which failed the p-value test above, the correlation coefficients between these variables are relatively low, so there is no multicollinearity between these variables.

### 4.3 Prediction Based on the Model

After fitting the linear regression model to the training data, the next step involves using the model to generate predicted values for both the training and test datasets. Then compare the actual values of the dependent variable with the values predicted by the model.

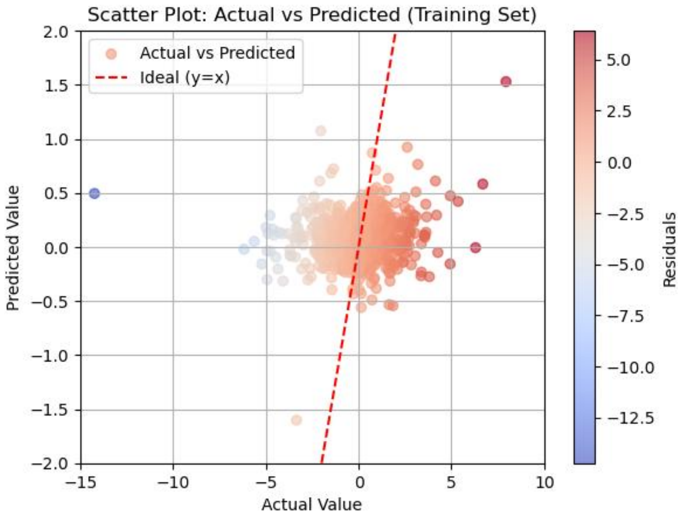


**Figure 2** Scatter Plot of Actual versus Predicted Data for the Training Set

To conclude, the model predictions perfectly match the actual values, the points on the scatter plot should align along a 45-degree diagonal line, indicating equality between actual and predicted values.

### 4.4 Model Evaluation

To evaluate the predictive performance of the multiple linear regression model, a systematic calculation of two key metrics, Adjusted $R^2$ and Root Mean Squared Error (RMSE), is conducted for both the training and test datasets.

$$Adj.R^2 = 1 - \frac{(1 - R^2)(1000 - 1)}{1000 - 6 - 1} \#(6)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{1000}(spy_t - spy_t)^2}{1000}} \#(7)$$

**Table 12** Assessment Table for the Regression Model

| | Train | Test |
|---|---|---|
| **Adj. $R^2$** | 0.0086 | 0.0130 |
| **RMSE** | 1.5456 | 3.5162 |

**Training Set Performance:**

The Adjusted $R^2$ for the training set is 0.0086, indicating that the model explains only 0.86% of the variance in the spy values after accounting for the number of predictors. The RMSE for the training set is approximately 1.546. This value reflects the average magnitude of prediction errors for the training dataset. While RMSE is relatively low, it should be interpreted in the context of the dependent variable's range to determine whether the errors are practically significant.

**Test Set Performance:**

The Adjusted $R^2$ for the test set is 0.0130, meaning that the model explains about 1.3% of the variance in the spy values for unseen data. While slightly higher than the training set $R^2$, this value remains very low. The RMSE for the test set is approximately 3.516, significantly higher than the training RMSE, which indicates that the model's predictions are less accurate when applied to new


Figure 3 Comparing Strategies for Training Data

data.

## 4.5 Strategy Building

The strategy begins by generating trading signals based on the predicted values from the linear regression model in training set. The signals are derived as follows:

- *A buy signal is generated if the predicted value (PredictedY) is positive.*
- *A sell signal is generated if the predicted value is negative.*

$$Profit_i = spy \times Order \quad \#(8)$$

Here, Order represents the directional trade signal (+1 or -1). This calculation captures the gain or loss from each trade based on the actual performance of *spy* (Hastie, Tibshirani & Friedman, 2009).

$$Total\ Profit_{training} = \sum Profit_i = \$116.49 \quad \#(9)$$

To evaluate the signal-based strategy, its cumulative returns are plotted against those of a buy-and-hold strategy, which assumes a constant long position throughout the training period.

From figure 3, there is a smooth upward trajectory in cumulative wealth reflects consistent profitability with limited volatility. Meanwhile, the points where the strategy diverges positively from the buy-and-hold benchmark highlight moments of particularly effective decision-making.

However, when turn to the Test set, the result shows less effective performance.

$$Total\ Profit_{test} = \sum Profit_i = \$71.01 \quad \#(10)$$

The total profit that the strategy made to Test data is smaller than that of Training set. When compared the strategy applied to Test data with buy-and-hold performance:

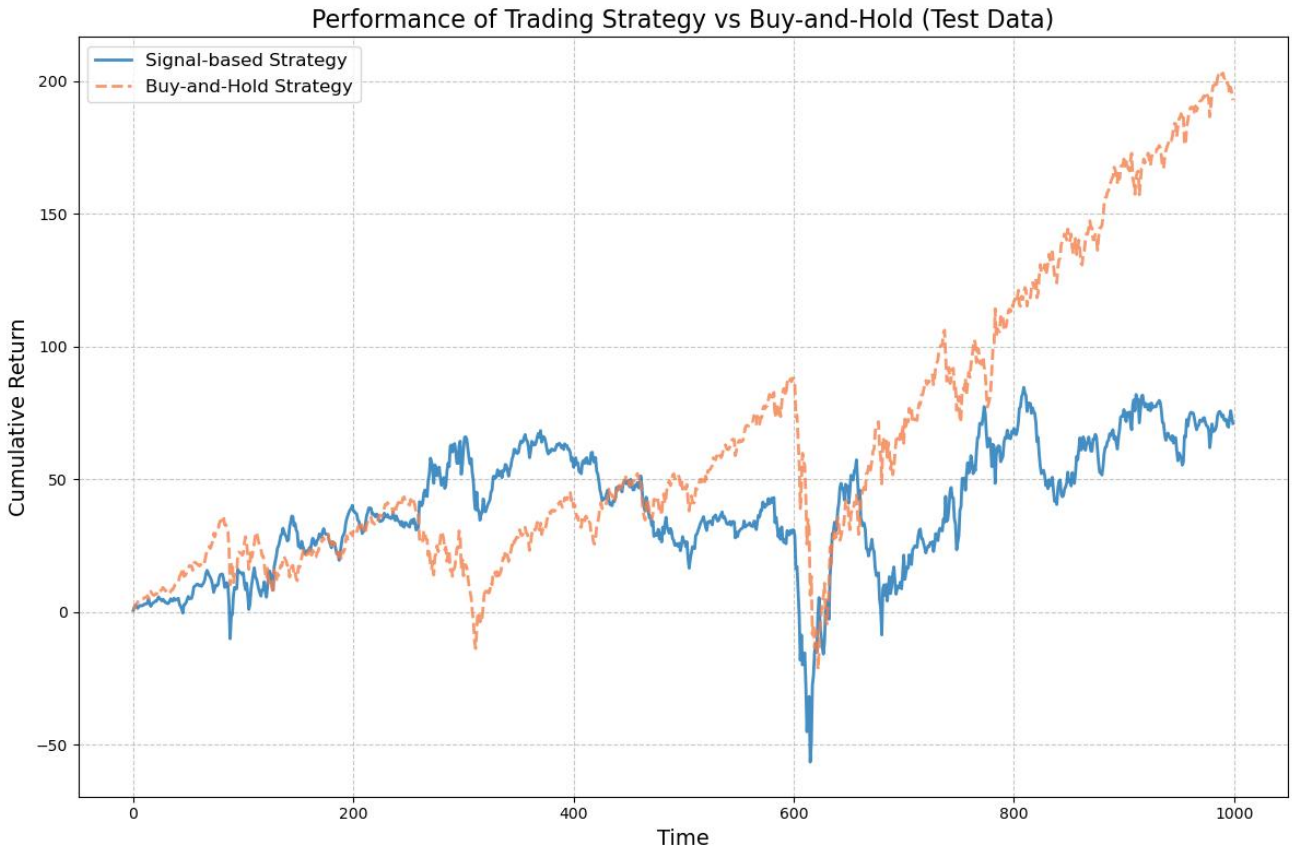


**Figure 4** Comparing Strategies for Test Data

The underperformance of the signal-based strategy compared to the buy-and-hold approach highlights the challenges of developing predictive models for financial markets. This outcome serves as a diagnostic tool, indicating areas where the model, feature selection, or strategy design may need refinement.

### 4.6 Evolution of Strategy

Evaluating the performance of a trading model requires robust metrics to quantify risk and return. Two widely recognized standards, the Sharpe Ratio and Maximum Drawdown, provide insight into the efficiency and stability of the strategy.

**Training Set Performance:**

$$Daily\ Sharpe\ Ratio_{training} = \frac{\mu_d}{\sigma_d} = 0.0790 \#(11)$$

$$Yearly\ Sharpe\ Ratio_{training} = Daily\ Sharpe\ Ratio_{training} \times \sqrt{252} = 1.2535 \#(12)$$

$$Maximum\ Drawdown_{training} = max\{\frac{Peak\ Wealth - Current\ Wealth}{Peak\ Wealth}\} = 0.0962 \#(13)$$

**Test Set Performance:**

$$Daily\ Sharpe\ Ratio_{test} = \frac{\mu_d}{\sigma_d} = 0.0186 \#(14)$$

$$Yearly\ Sharpe\ Ratio_{test} = Daily\ Sharpe\ Ratio_{test} \times \sqrt{252} = 0.2958 \#(15)$$

$$Maximum\ Drawdown_{test} = max\left\{\frac{Peak\ Wealth - Current\ Wealth}{Peak\ Wealth}\right\} = 0.3925 \#(16)$$

To conclude:

**Table 13** Yearly Sharpe Ratio and Maximum Drawdown for Strategy

| | **Train** | **Test** |
|---|---|---|
| Yearly Sharpe Ratio | 1.2535 | 0.2958 |
| Maximum Drawdown | 0.0962 | 0.3925 |

The large disparity between the training and test Sharpe Ratios reflects a potential overfitting issue. The model performs well during training but fails to maintain consistent risk-adjusted returns when applied to new data. Furthermore, the higher Maximum Drawdown on the test set highlights the model's inability to mitigate large losses in an out-of-sample environment, undermining its practical utility in real-world trading scenarios.

## 5. Findings and Analysis

### 5.1. Model Performance and Interpretation

The low $R^2$ (0.015) and adjusted $R^2$ (0.009) values indicate that the predictors collectively explain only a small fraction of the variance in SPY. This underscores the complexity of SPY price movements and the influence of unmodeled factors (Tsay, 2010).

Statistically insignificant coefficients for DAX, All Ordinaries, and Nikkei imply limited direct impact from these indices. This aligns with their weak correlation coefficients observed during EDA.

Significant deviations from normality and high kurtosis in residuals point to model inadequacies in capturing extreme movements or non-linear relationships in the data.

### 5.2. Strategy Effectiveness Analysis

The strategy outperformed the buy-and-hold approach, yielding higher cumulative returns and a favorable Sharpe Ratio of 1.2535. This reflects the model's ability to exploit historical patterns in a controlled environment (Montgomery, Peck & Vining, 2012).

The strategy underperformed relative to buy-and-hold, producing a lower Sharpe Ratio (0.2958) and higher maximum drawdown (0.3925). This indicates reduced robustness in capturing patterns within unseen data, highlighting overfitting and the need for more generalizable models.

### 5.3. Analysis of Model Limitations

The inability to capture substantial variance in SPY suggests that the linear model may not fully account for the dynamic and complex nature of financial markets. Factors like macroeconomic indicators or non-linear relationships may need inclusion.

Furthermore, the disparity between training and test performance metrics underscores overfitting, with the model tailoring to training data patterns at the expense of generalization (Tsay, 2010).

Although multicollinearity tests did not indicate significant issues, the inherent high correlations between U.S. indices (SP500 and DJI) may have contributed to redundancy in predictor variables.

### 5.4. Recommendations for Improvement

**(1) Enhance Feature Selection**

The model relies primarily on lagged SPY values and contemporaneous global indices, which explain only a small fraction of SPY's variance. This suggests the need for additional predictors to capture overlooked drivers of SPY price movements. For example, by adding predictors such as Nasdaq in US market, cac40 in French market, and Hsi in Hong Kong market, the model can achieve better performance:

$$\begin{aligned} spy_t &= \beta_0 + \beta_1 spy_{lag1_t} + \beta_2 sp500_t + \beta_3 dji_t + \beta_4 daxi_t \\ &+ \beta_5 aord_t + \beta_6 nikkei_t + \beta_7 nasdaq_t + \beta_8 cac40_t + \beta_9 hsi_t + \varepsilon_t \end{aligned} \#(17)$$

**(2) Adopt Advanced Predictive Models**

The linear regression model oversimplifies the relationship between SPY and global indices, failing to capture non-linear dynamics or complex interactions. Transition to algorithms capable of modeling non-linear relationships and interactions among predictors. Machine learning models such as Gradient Boosting Machines (GBM), Random Forest, and neural networks can better capture complex dependencies and provide improved predictive accuracy. Additionally, ensemble methods that combine multiple model outputs can enhance generalization and reduce the risk of model bias. Consider implementing XGBoost or LightGBM for faster computation and superior handling of large datasets. Regularly evaluate model performance using metrics like RMSE (Root Mean Square Error) and $R^2$ to monitor improvements over the baseline linear model.

**(3) Address Multicollinearity and Feature Redundancy**

High correlations among U.S. indices (e.g., SP500 and DJI) may introduce redundancy and reduce model interpretability. Apply Ridge or Lasso regression to shrink coefficients of less relevant predictors and reduce the impact of multicollinearity. Ridge regression can help mitigate overfitting by penalizing large coefficients, while Lasso regression can perform feature selection by shrinking some coefficients to zero. Conduct a Variance Inflation Factor (VIF) analysis to identify and quantify multicollinearity among features. If severe multicollinearity persists, consider applying Principal Component Analysis (PCA) to transform correlated variables into orthogonal components, thereby improving model stability and reducing overfitting.

**(4) Improve Model Validation and Evaluation**

The model's performance deteriorates significantly on test data, indicating overfitting to training data patterns. Replace a single train-test split with k-fold cross-validation to ensure the model’s robustness across different subsets of data. Implementing a time-series cross-validation approach, such as walk-forward validation, is particularly important for financial data to preserve the temporal order and simulate real-world forecasting scenarios. Evaluate model performance across multiple folds using statistical tests to verify consistency. In addition to conventional error metrics (e.g., MAE, RMSE), track directional accuracy to assess how well the model predicts market trends. Incorporate backtesting to evaluate the model’s profitability under realistic trading conditions, considering transaction costs and slippage.

### （5） Refine Trading Strategy

The signal-based trading strategy underperformed the buy-and-hold approach in the test set, indicating inefficiencies (Brooks, 2014). Instead of a binary signal (buy/sell), use thresholds to differentiate between strong and weak signals, adjusting trade sizes accordingly. For example, generate buy signals only if the predicted price change exceeds a 1% threshold to account for transaction costs and noise. Implement a dynamic position-sizing strategy where trade size increases with stronger predictive confidence. Additionally, consider using stop-loss and take-profit mechanisms to manage risk and lock in profits. Analyze the Sharpe ratio and maximum drawdown to assess and optimize the risk-return tradeoff of the refined strategy. Explore alternative signal-generation techniques, such as momentum-based indicators (e.g., Moving Average Convergence Divergence) or volatility-adjusted signals, to improve decision-making.

### （6） Feature Engineering for Temporal Effects

The model does not explicitly account for temporal market patterns or seasonality. Extend lagged features beyond SPY (e.g., lagged values for other indices like SP500 and Nikkei) to capture delayed effects. For example, add a 3-day lag for SP500 and DJI to investigate their cumulative impact on SPY, or add a binary feature indicating whether the day is a Monday, accounting for potential “Monday Effect” anomalies in stock prices. Introduce rolling-window features (e.g., 5-day and 20-day moving averages) to capture short-term and long-term market momentum. Consider encoding macroeconomic events or earnings announcement dates as categorical variables to account for external shocks. Furthermore, investigate seasonality by including quarterly and yearly dummies to capture cyclical market behaviors, such as the "January Effect" or fiscal quarter-end rebalancing.

## 6. Conclusion

This study investigated a signal-based trading strategy for the SPY ETF by incorporating multiple linear regression and examining the relationships between various global indices. The analysis revealed that while U.S. indices exhibited strong correlations with SPY, the model's overall explanatory power was significantly limited, with very low adjusted $R^2$ values of 0.009 and 0.013 for the training and test data, respectively. These results indicate that the model was not able to capture much of the variance in the SPY ETF's price movement, limiting its practical utility. Although the trading strategy showed promising results on the training data, it performed poorly on the test set, a clear indication of overfitting. This suggests that the model had learned specific patterns in the training data that did not generalize well to unseen data, highlighting a fundamental flaw in its predictive power.

Several limitations of the study were identified, including the reliance on a linear model, which likely oversimplifies the complexities of the financial markets. Financial markets are influenced by a multitude of factors, such as macroeconomic trends, geopolitical events, and market sentiment, which a linear model fails to capture. Additionally, the exclusion of volatility measures or other macroeconomic indicators, which are often crucial in understanding market dynamics, further hindered the model's performance.

For future research, it is recommended to explore more advanced, non-linear models, such as machine learning techniques, which have shown greater promise in identifying complex, non-linear relationships in financial data. Expanding the feature set to include volatility indices, sentiment analysis, and macroeconomic data could also provide a more comprehensive view of the market dynamics that affect the SPY ETF. By refining the trading strategy and improving model robustness, future studies could make significant strides in enhancing profitability and developing trading strategies that are more adaptable to real-world market conditions. This would represent a meaningful step forward in the development of ETF-based trading methodologies with broader applicability and potential for real-world profitability.

**References:**

Bollapragada, R., Savin, I. and Kerbache, L. (2013) 'Price forecasting and analysis of exchange-traded fund', *Journal of Mathematical Finance*, 2013, pp. 181–191. Available at: https://www.scirp.org/journal/paperinformation?paperid=29338 (Accessed: 17 December 2024).

Brooks, C. (2014) *Introductory Econometrics for Finance*. 3rd edn. Cambridge: Cambridge University Press.

Dhanaraj, S., Gopalaswamy, A. and Sureshbabu, M. (2013) 'Dynamic interdependence between US and Asian markets: an empirical study', *Journal of Financial Economic Policy*, 5(3), pp. 220–237.

Dias, R., Galvão, R., Irfan, M., Alexandre, P., Gonçalves, S. and Almeida, L. (2024) 'Delving into the exchange-traded funds (ETFs) market: understanding market efficiency', *Journal of Ecohumanism*, 2024.

González-Mancha, J., Frausto Solís, J., Mora Guadalupe Castilla, V.G. and Terán Villanueva, D. (2017) 'Financial time series forecasting using simulated annealing and support vector regression', *International Journal of Combinatorial Optimization Problems and Informatics*, 8, pp. 10–18.

Hastie, T., Tibshirani, R. and Friedman, J. (2009) *The elements of statistical learning: Data mining, inference, and prediction. 2nd edn*. New York: Springer.

Hill, J.M., Nadig, D. and Hougan, M. (2015) A comprehensive guide to exchange-traded funds (ETFs).

James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013) *An Introduction to Statistical Learning*: *With Applications in R*. New York: Springer.

Jamil, I., Kogid, M., Lim, T. and Lily, J. (2023) 'Pre- and Post-COVID-19: The impact of US, UK, and European stock markets on ASEAN-5 stock markets', *International Journal of Financial Studies*.

Liu, Y. (2024) 'A study on the effectiveness of SPY ETF prediction based on signal trading', *Academic Journal of Business & Management*, 6(7), pp. 1–12. Available at: https://francis-press.com/papers/16369 (Accessed: 17 June 2024).

Marszk, A. and Lechman, E. (2020) 'Application of diffusion models in the analysis of financial markets: evidence on exchange-traded funds in Europe', *Risks*.

Montgomery, D.C., Peck, E.A. and Vining, G.G. (2012) *Introduction to Linear Regression Analysis*. 5th edn. New York: Wiley.

Pan, Y., Hu, Y., Haran, E. and Manago, S.M. (2014) 'Co-movement of European stock markets based on association rule mining', *Proceedings of the International Conference on Data Mining*.

Sakthivel, P. and Kamaiah, B. (2012) 'Interlinkages among Asian, European and the U.S stock markets: a multivariate cointegration analysis', *Journal of Economics and Behavioral Studies*, 4(3), pp. 129–141.

Sermpinis, G., Verousis, T., and Theofilatos, K. (2016) Adaptive Evolutionary Neural Networks for Forecasting and Trading without a Data-Snooping Bias. *J. Forecast.*, 35: 1–12. doi: 10.1002/for.2338.

Shamsuddin, A. and Kim, J.H. (2003) 'Integration and interdependence of stock and foreign exchange markets: an Australian perspective', *Journal of International Financial Markets, Institutions and Money,* 13(3), pp. 237–254.

Tsay, R.S. (2010) *Analysis of Financial Time Series*. 3rd edn. New York: Wiley.

Vasylieva, O., Maksyshko, N. and Pedan, D. (2020) 'Comparative analysis of quality to forecast models of exchange traded funds dynamics', *Scientific Bulletin of Uzhhorod National University*, 4(48), pp. 141–148.